

Comment on Broadband Mapping

Broadband mapping is a term in high currency right now, and could become an important part of the Broadband stimulus initiative. While it is likely to have many benefits, these benefits will be achieved only if the process is conducted with careful attention to quality. As someone who has spent 28+ years doing social research, I am concerned to observe that much of the recent broadband mapping is not of high quality, to the extent that to make decisions based on it “results” would be irresponsible and ill advised.

Since the documents from the Federal Register request that commenters focus on the process (AKA methodology) for broadband mapping, that is the topic of this comment. However, there is a lot to say above and beyond process and methodology, which I shall address in separate comments.

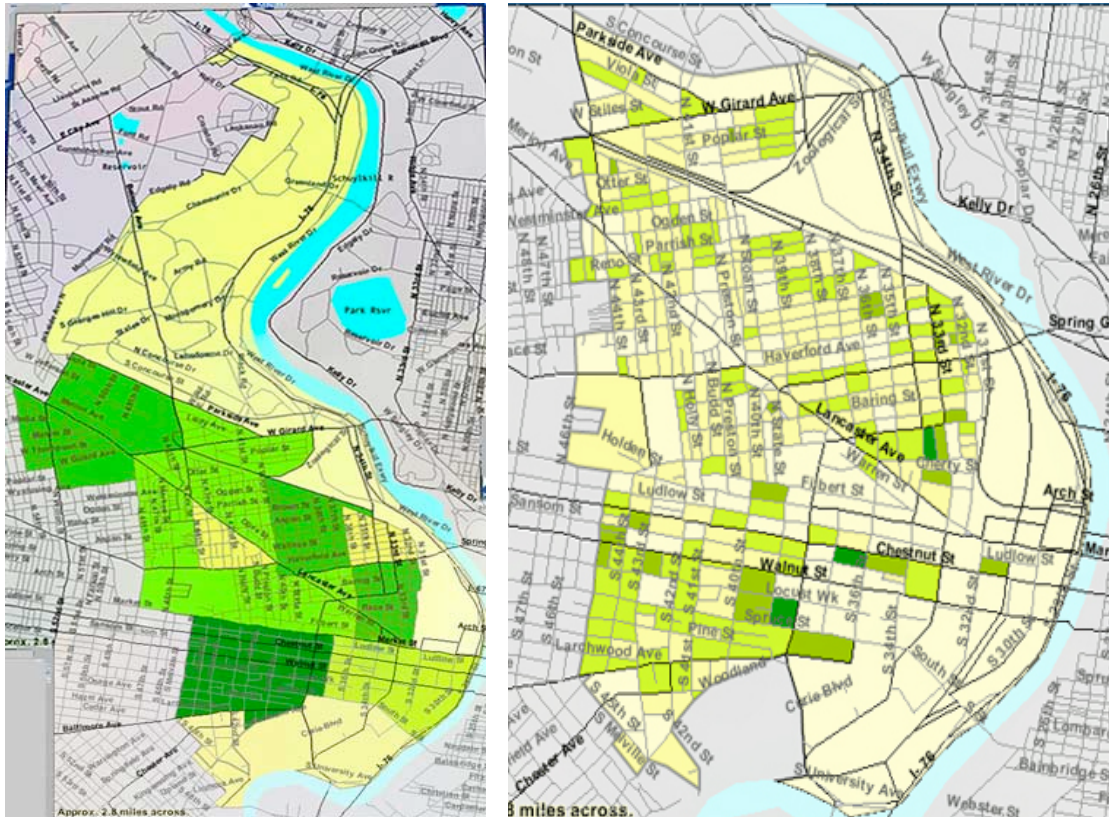
The Mapping Process: Recommendations

- Before designing the research or collecting a single datum, the team must come to agreement on a small set of explicit goals for the project. Questions they must address include: “Why are we doing this?,” “How will we know whether the effort has succeeded or failed?,” “What can we do to enhance the likelihood of success?,” “Assuming success, what will we do with the information that results?,” and finally, “How will this effort contribute to the creation of jobs?” Note that job creation may have indirect links, going beyond those created with the specific intent of executing the mapping process. However, the jobs question must be addressed.
- The mapping research should begin with a thorough review of the literature. This step is essential since it mitigates the danger of overlooking unanticipated requirements, re-inventing the wheel, or taking approaches that have already been proven not to work. It includes reviewing secondary sources (e.g. published research, analyses, etc.), conducting interviews with experts on demography, mapping, social research among difficult-to-access populations, etc. as well as interviews with people living in areas purportedly enjoying good access, intermittent- to-poor access, and no access. Service providers should also be included in this “hypothesis generation” phase of the process. This stage is essential to ensure that one is asking the right questions, and that mapping researchers can correct for distortive factors as they design and implement the mapping exercise. Moreover, it is essential to establish a finite number of hypotheses to be tested, the NTIA will benefit from clarity about the purpose and uses of the data, as well as lessons learned in which they may place their confidence.
- In conducting the study, it is imperative that data be collected from multiple sources, and that among them is a sample of people living and working in the areas under study. There is simply no substitute for this. I have seen studies conducted using only data supplied by operators, after changing items unnamed to “Preserve provider competitive advantage and confidentiality”¹ The lack of data from any other source, the lack of specificity about what data may have been excluded, and the lack of any

¹ Rachelle Chong, California Public Utilities Commission: Mapping Broadband: California’s Story NARUC Summer Meeting Portland, OR, July 2008

human subjects in this mapping exercise, together render the resultant information highly unreliable as a basis for decision making.

- Two interrelated factors of extreme import, both overlooked in most of the recent mapping exercises, are population size and population density. As the reader is probably aware, the US census collects and can report on geographic units as small as the census block². The reason for the importance of the choice of unit of measurement is that in urban areas, there are likely to be several distinct neighborhoods with distinct, different socio-economic characteristics, within a single zip code. This is far less likely to be true in rural areas. For example, in Philadelphia, zip code 19104, there were, as of the 2000 census, 50,125 residents in a 3.02 square mile area. Meanwhile, zip code 19104 had more residents than did 25 of the Pennsylvania's 67 counties, which together cover an area of 16,001.49 square miles³
- Rather than describing why this is important, I'll illustrate with two different views of zip code 19104. In both cases the darker colors indicate higher population density. The first (on the left) shows population density for zip code 19104 at the Census tract



level, where the median population is “smoothed” to reflect the larger unit of measurement. The second shows the same zip code at the block level.

² The “Census Block,” a subdivision of a census tract is the smallest geographic unit for which the Census Bureau tabulates 100-percent data. Many blocks correspond to individual city blocks bounded by streets, but blocks - especially in rural areas - may include many square miles and may have some boundaries that are not streets.

³ Source: US Census Bureau, 2000 Census

The two pictures look very different, while in fact the only difference is that the second view is more granular, showing that there are a few densely populated areas, and quite a few that are not densely populated at all. In fact, one area with very low population density is part of Fairmount Park, the largest urban park in the nation. No wonder it also has a low median income: nobody lives there!

- The ‘pictures’ of 19104 are similarly different when one looks at household income, and – for that matter – at broadband availability. In 19104 (AKA West Philadelphia), there are some small pockets of wealth interspersed with much poorer areas in the areas close to the University of Pennsylvania, which has offered faculty very advantageous mortgage loans to persuade them to live near its campus. Both the University of Pennsylvania and Drexel University are located in a section of this zip code called “University City.” Both Universities have high-speed networks wired and wireless, but they are only to staff, students and faculty. These networks are password protected, which means that local inhabitants who do not have any connection with either University, and/or do not have the means to purchase services privately, remain unserved.
- This problem can be seen on a larger scale as well. In California, 93.3% of the state lives in urban areas, according to the Census Bureau. Here, population size and density are such that analysis on the Census tract or zip code level is extremely likely to obscure reality. For instance, San Diego County has more inhabitants than do the states of Alaska, North Dakota, Vermont, the District of Columbia and Wyoming—combined. Urban areas are characterized by extremes of wealth and poverty in extreme proximity to one another. It is very common for certain wealthy individuals, as well as for businesses, to be able to buy broadband access, and for the people walking on the street below to have no access. If those conducting Broadband mapping have little knowledge of demographics, best practices, and/or of the many nuances involved in social research, we run the risk of getting a mapping process that produces distorted and distortive results.
- The research team must also have strong expertise in data analysis, since many of Census statistics are based on projections with numerous inputs, complex predictive equations, and interpretive caveats in the form of “confidence’ intervals” that can have a large impact on the usability of the data.
- Optimally, the Broadband Mapping initiative should be a socially engaging process, one that involves communities and serves as a means of encouraging adoption. To be as accurate as possible, the process should be iterative. In other words, mappers should collect data, analyze it, then and then go back to those in the mapped areas to find out if they got it ‘right.’ Community involvement should be encouraged right from the start, because it is only through broad involvement that the process will take on meaning and value for its most important stakeholders, the American people.